Analytics Vidhya

Harnessing ML and NLP for Elevated Customer Experiences

The DataHour

Seth Levine

Lead ML Scientist, Loris

March 2024

About Me

- Lead ML Scientist at Loris
- Host of Learning from Machine Learning Podcast
- M.S. Computer Science
- B.A.S. Complex Human Behavior and Evolution

Agenda

- What is Loris?
- What ML & NLP Challenges do we face?
- Topic Modeling using BERTopic
- Text Classification with SetFit
- Evaluating ML Models
- Demo
- Takeaways
- Q&A

Loris

What is Loris?

Loris

Company Highlights

Clients Include:

FINTECH	flex. bill zip
TECH	fiverr slice
TELECOM	mintmobile ultra mobile
ECOMM	THESIS BARK BOX
BPOs	horatio redial

Investors Include:

FLOODGATE

vertex VENTURES

homebrew


Jeff Weiner
(LinkedIn)

bowcapital

servicenow

NYC, TLV
Offices

2018
Founding

>200M
interactions analyzed

6 continents
Global

Loris

QA (1%)

The Problem

Every company's greatest asset is its most underutilized

Billions of customer interactions contain a goldmine of untouched insights

Surveys (9%)

The Solution

We use AI to transform every customer interaction into an opportunity for insights, optimization and growth

Loris' Approach

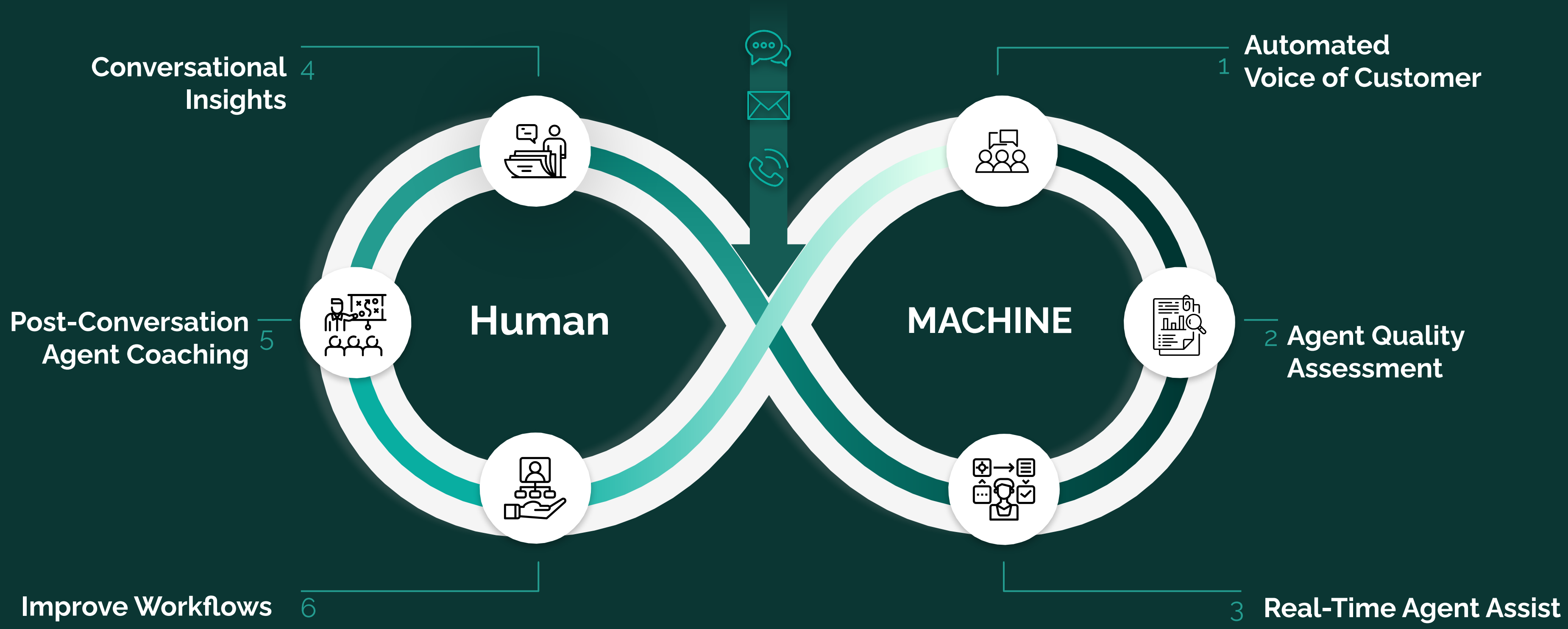
Insights

- The challenge is modeling a domain to solve its specific problems
- Foundation models are becoming a commodity (both commercial and OS)
- Understanding the domain and gathering the right data is a differentiator

Approach

- Model the domain of customer support conversations (conversation phases, dialog acts, etc.)
- Use proprietary data to optimize models for customer support (e.g., Customer Sentiment, Conversation Markers)
- Develop customized Contact Driver models for clients
- Enable a feedback loop on how users interact with our product and predictions

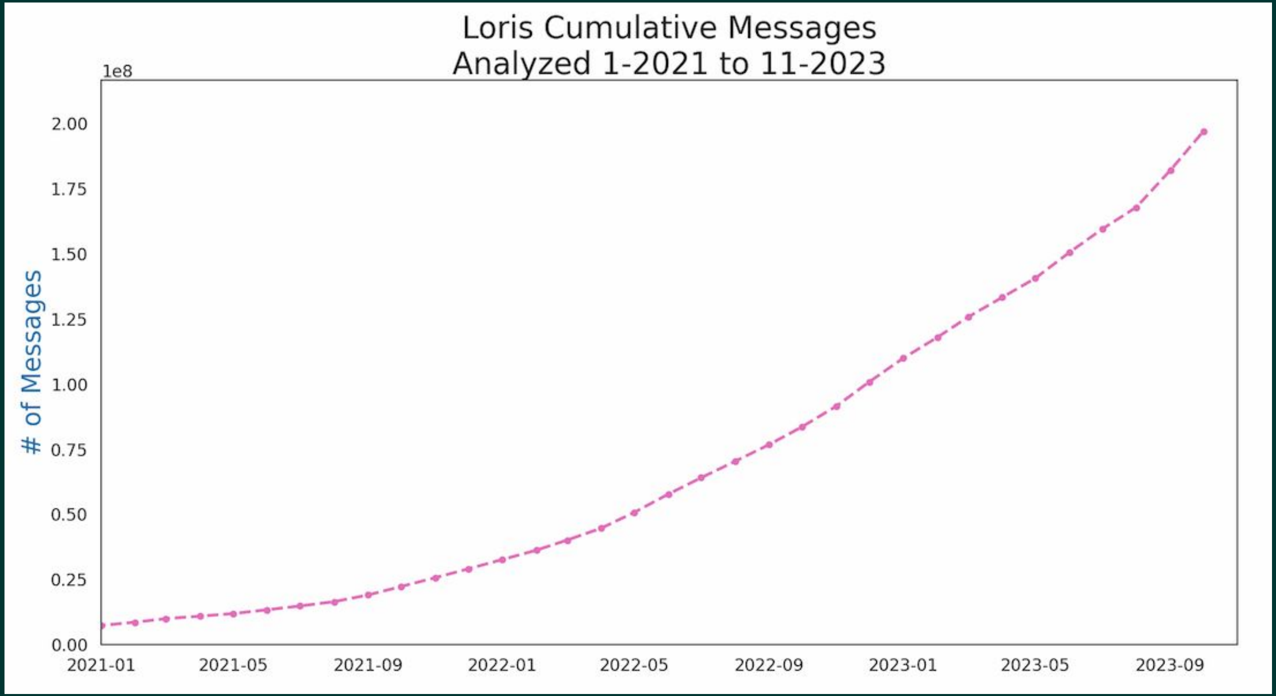
Loris Combines the Best of Humans ∞ MACHINE



Improve customer experience, operational efficiency, and insight

Loris

Our Platform is Scaling Fast



Over 200 million customer care messages analyzed
Analyzing over 5 million messages per month

Loris

Omni-Channel

Chats

Emails

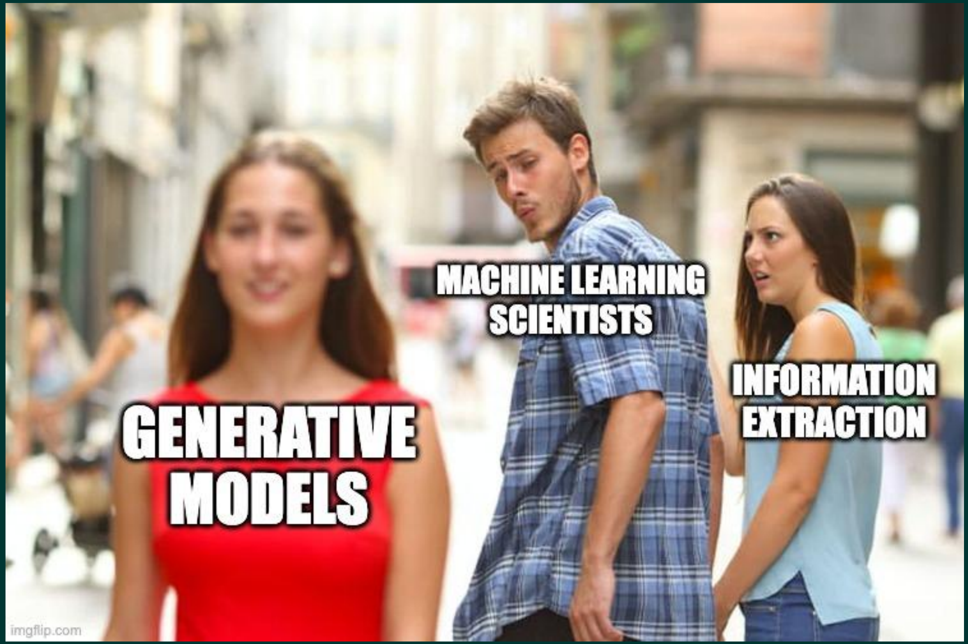
Messaging Apps

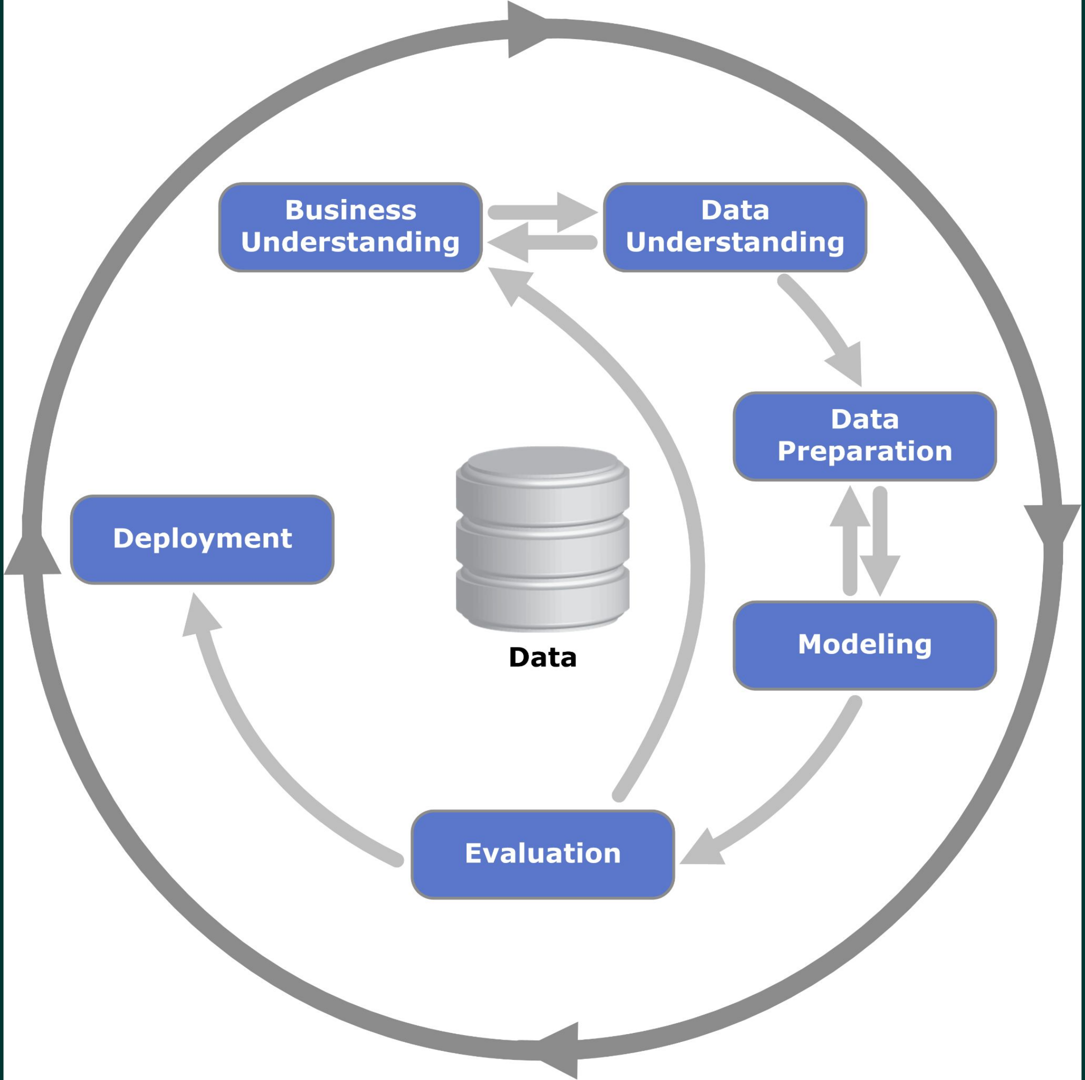
Phone calls

State of Machine Learning

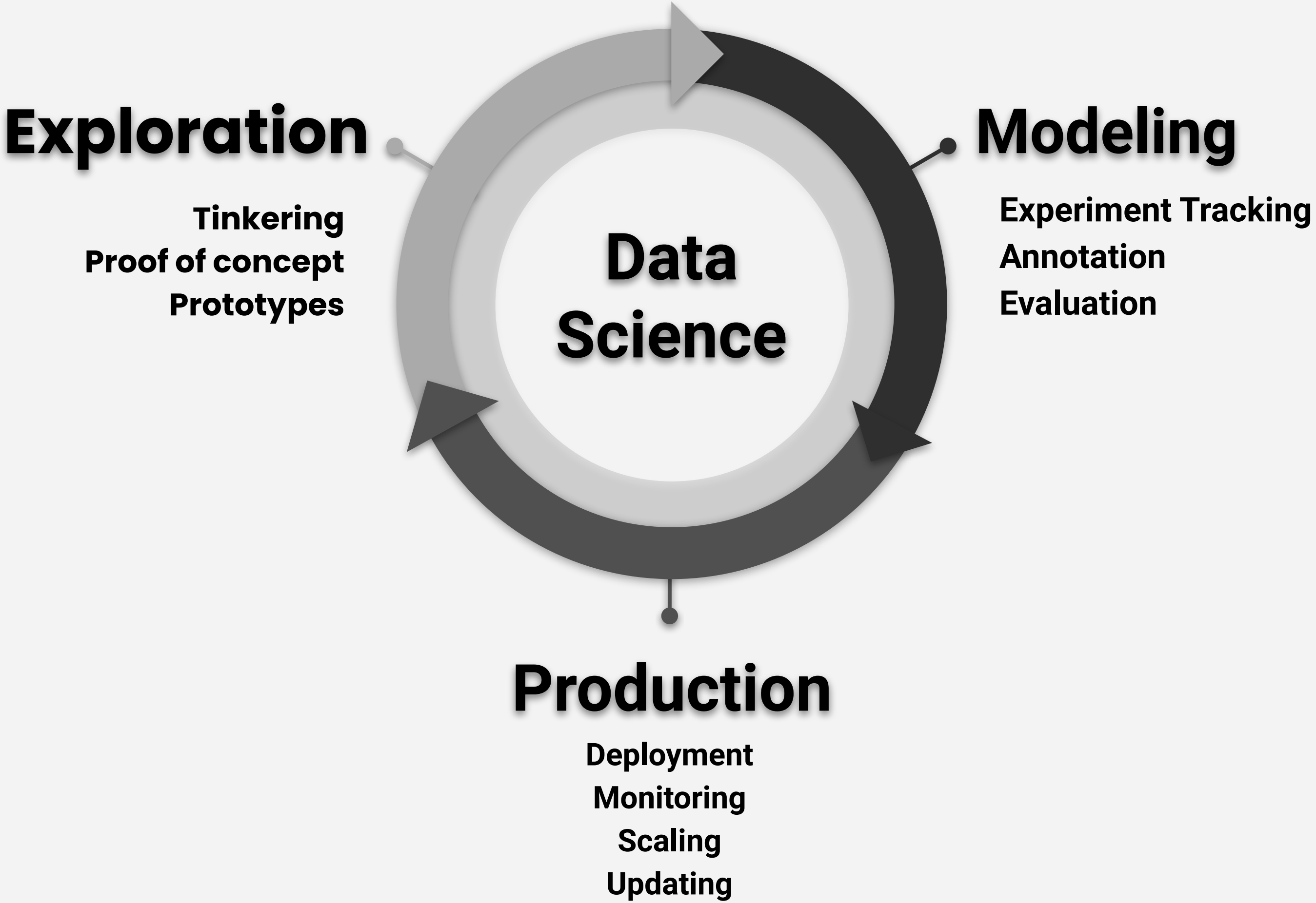
Loris

How have ChatGPT-like models changed our approach?





[Cross-industry standard process for data mining - Wikipedia]


Exploration
Tinkering
Proof of concept
Prototypes
Data Science
Modeling
Experiment Tracking
Annotation
Evaluation
Production
Deployment
Monitoring
Scaling
Updating

Machine Learning and NLP at Loris

Loris

Main Questions

How is the customer feeling?

+

Why did the customer contact the company?

+

How did the conversation go?

Conversation Modeling and Intelligence

Loris Generalized Models

Message-Level

- Sentiment Model
- Sentiment Target
- Conversation Markers
- Profanity
- Polite Refusal
- Question Detection
- ...

Conversation-Level

- Conversation Quality (CQ)
- Summarizer
- Resolution
- Agent Actions

Loris

Loris Custom Client Models

- Client-specific text classification models that each client can use to understand their customers' reason for contact (or intent)

Loris

Underlying Need

What problem are we trying to solve?

- Identify what the conversation is about
- Understand a company's contact drivers

Knowing the contact driver of a conversation allows Loris and our clients to segment conversations into meaningful groups for improvement opportunities

The ability to interpret what a customer is looking for no matter how they word the request

Company can analyze trends, route, create and improve workflows

Intent Frequency Over Time

Loris

Intent Frequency Over Time

Loris

Evaluation

Coverage: % of conversations that are assigned 1 or more intent(s)

Precision: Fraction of detected intents that are correctly predicted

- When an intent is predicted, is it correct?
- Examined on the class-level

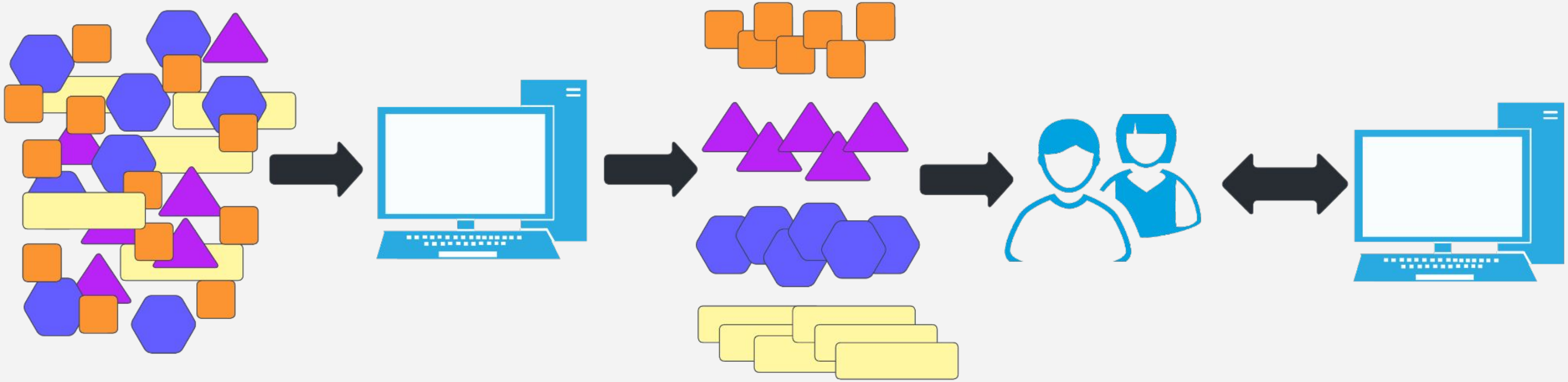
Frequency: In historical data, does the frequency of each contact driver make sense to clients?

- Proxy for recall on the class-level

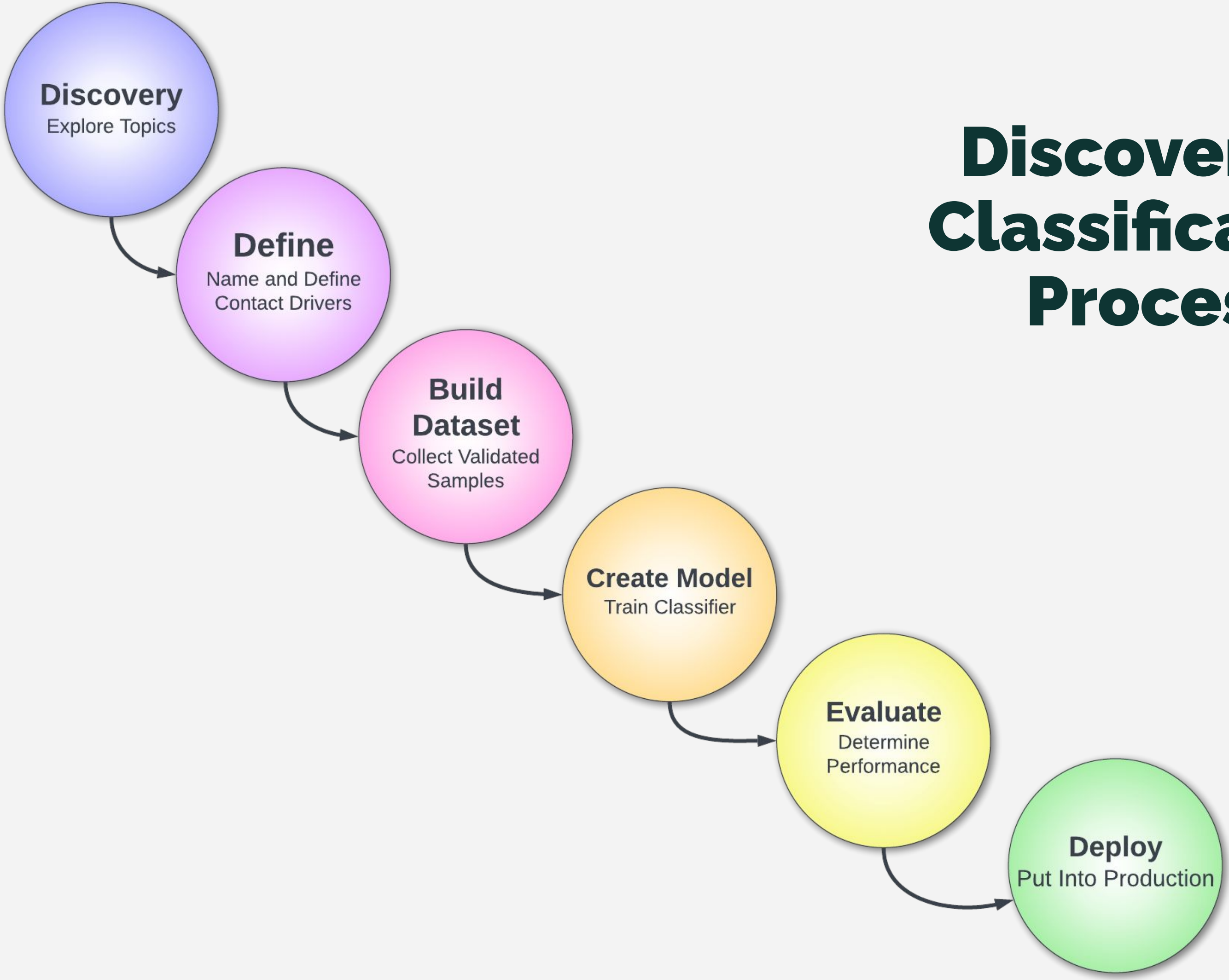
Loris

Topic Modeling supports Text Classification

Unsupervised Learning supports Supervised Learning



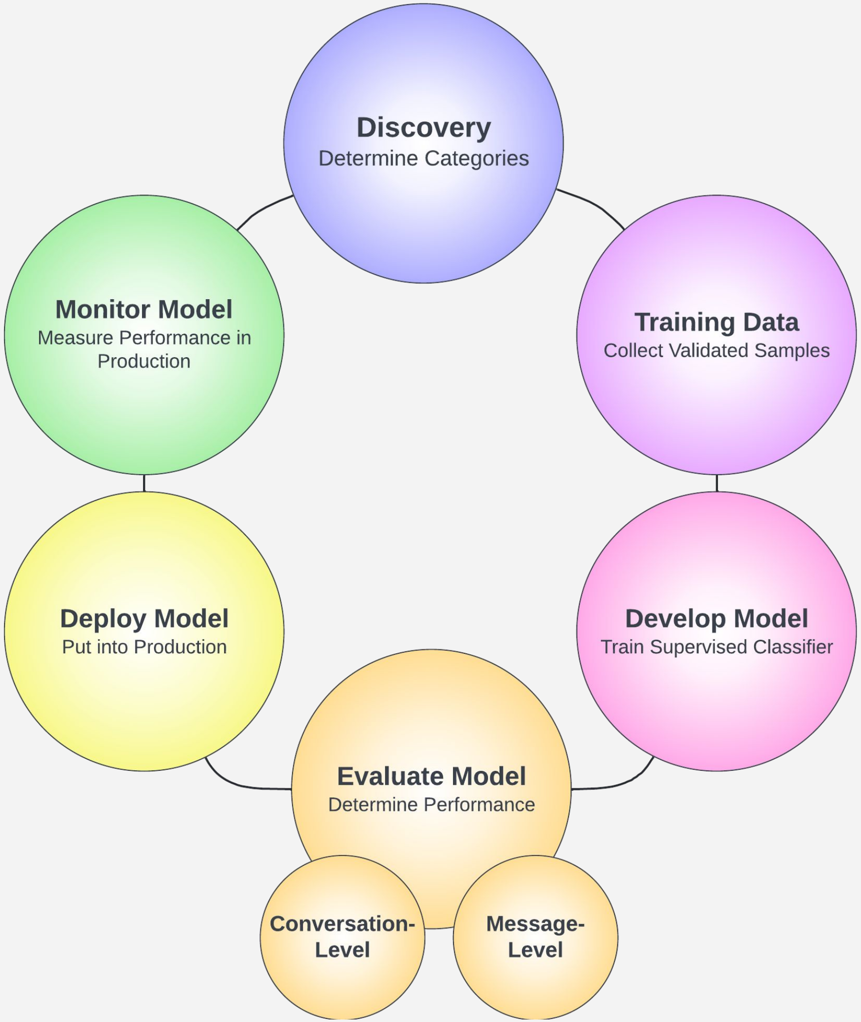
Loris



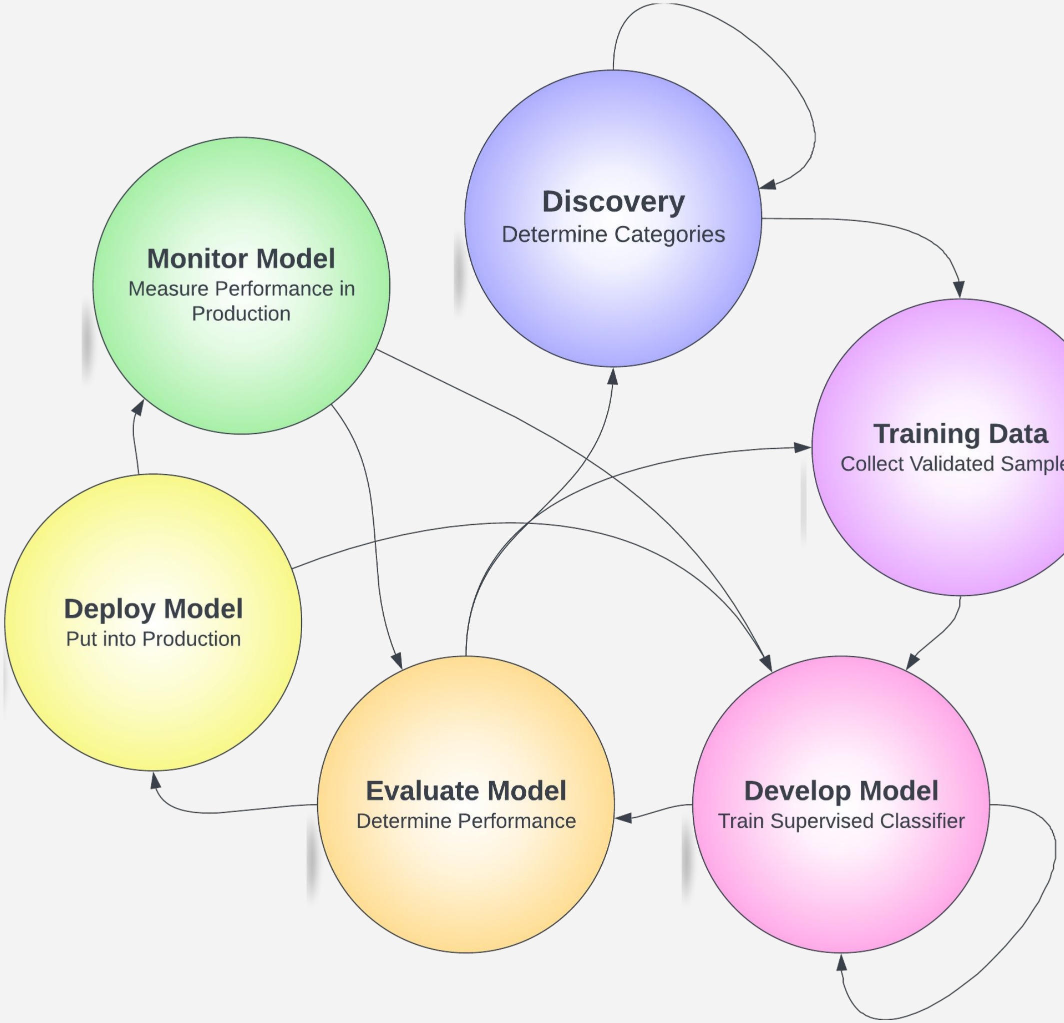
Discovery & Classification Process

Loris

Discovery and Classification Process

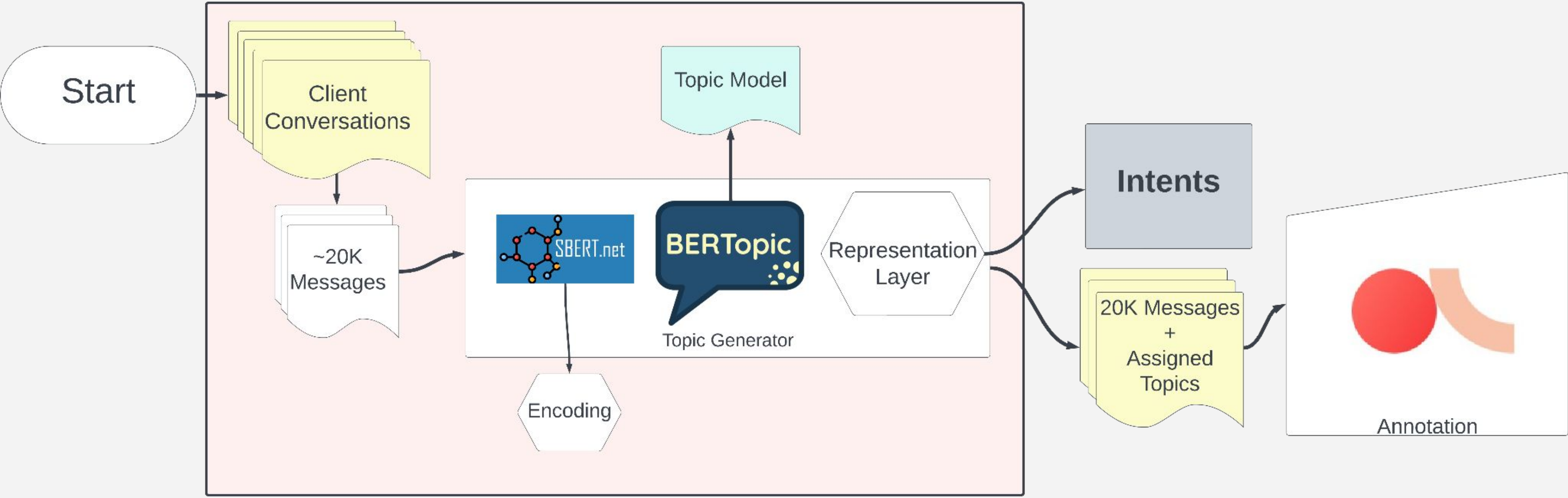


Discovery and Classification Process



Loris

Discovery



Loris

Topic Modeling using BERTopic

Input Text → Sentence Representation → Dimensionality Reduction → Clustering → Cluster Naming → Topics

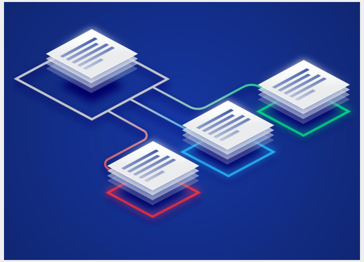
Unlabeled Messages

SBERT.net
Sentence-Transformers

UNIFORM MANIFOLD
UMAP
APPROXIMATION & PROJECTION

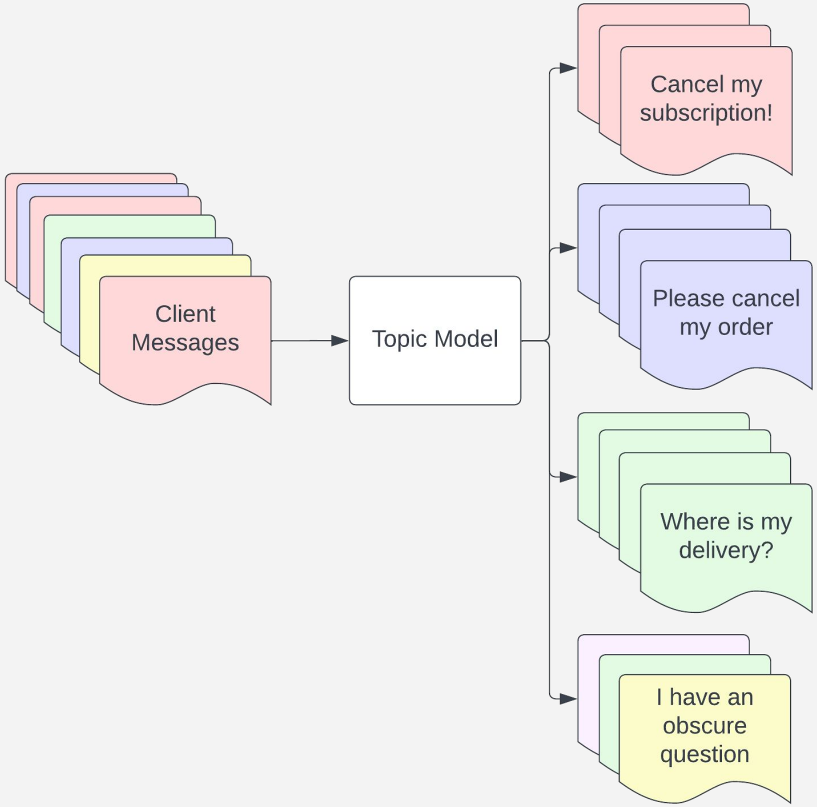
HDBSCAN

Azure
OpenAI



Discovery

- Topic Modeling
- Use unsupervised learning to find structure in data
- Create clusters and groups of messages
- Leverage many techniques
  - Unsupervised learning
  - Seeded topic modeling
  - Multi-aspect representation
- Generate Names for Topics
- Generate Definitions for Topics
- Output representative samples for each topic



Documents and Topics

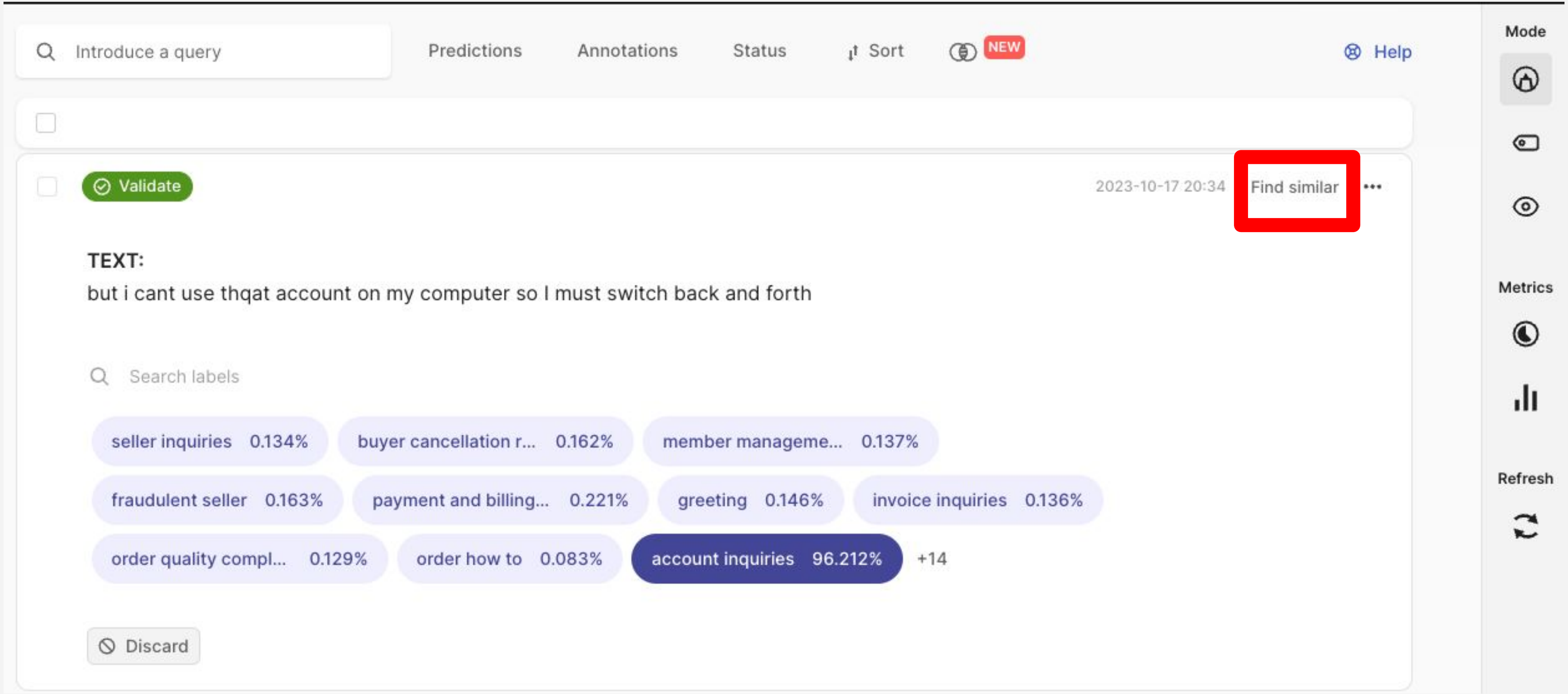
Discovery Outputs

- Number of classes
- Generated names of classes
- Generated definitions of classes
- Representative samples for each class

Gather Labeled Samples

- Validate samples in Argilla
- Find a small sample for each class
- SetFit only needs a few samples to train
- We need additional samples to validate model
- Use similar to build up dataset quickly

Label Samples Using Argilla

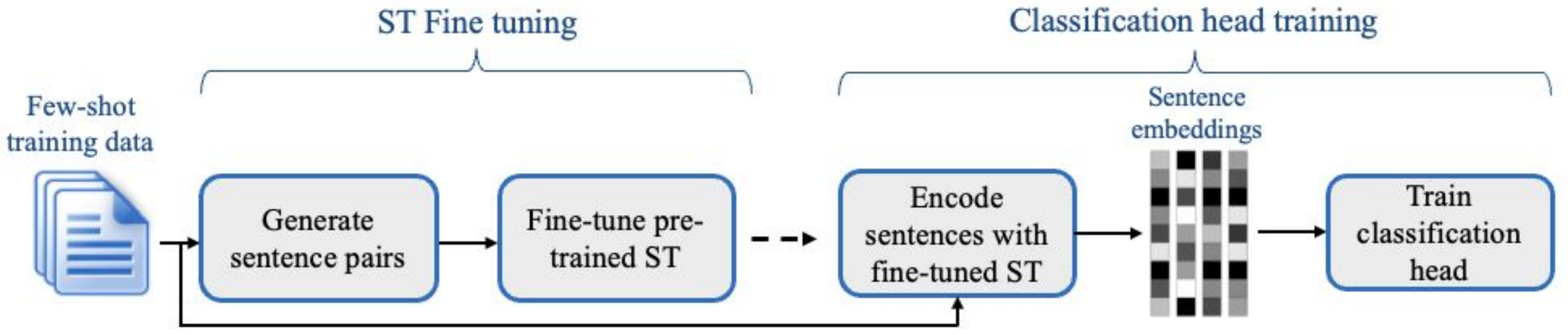


SetFit

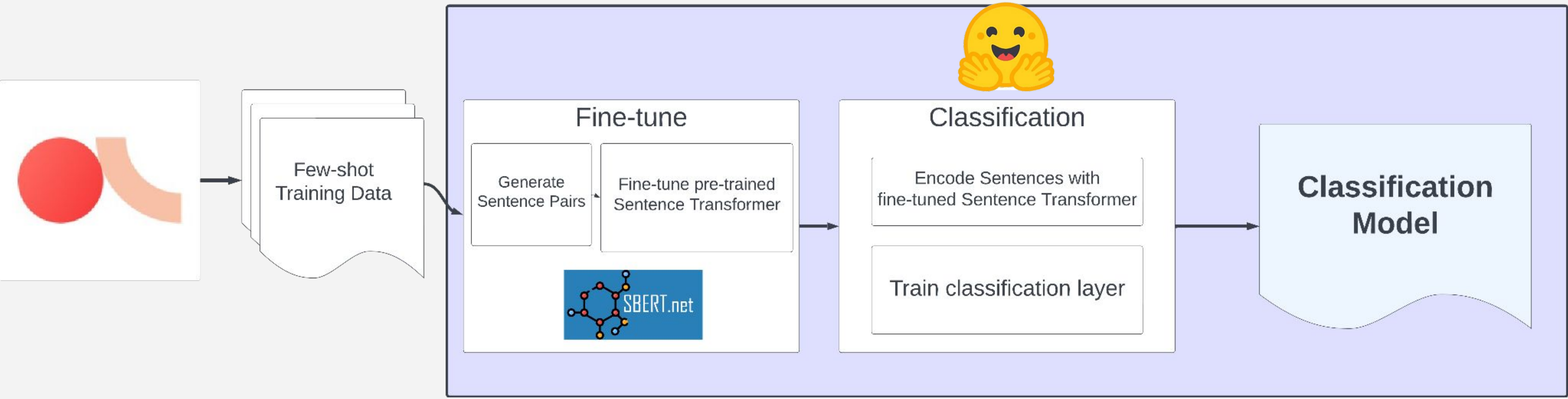
Loris

What is SetFit?

- Efficient Few-shot Learning with Sentence Transformers
- SetFit on HuggingFace
- GitHub - huggingface/setfit
- Used for multi-class text classification
- [arXiv] Efficient Few-Shot Learning Without Prompts
- Collaboration between HuggingFace and Intel



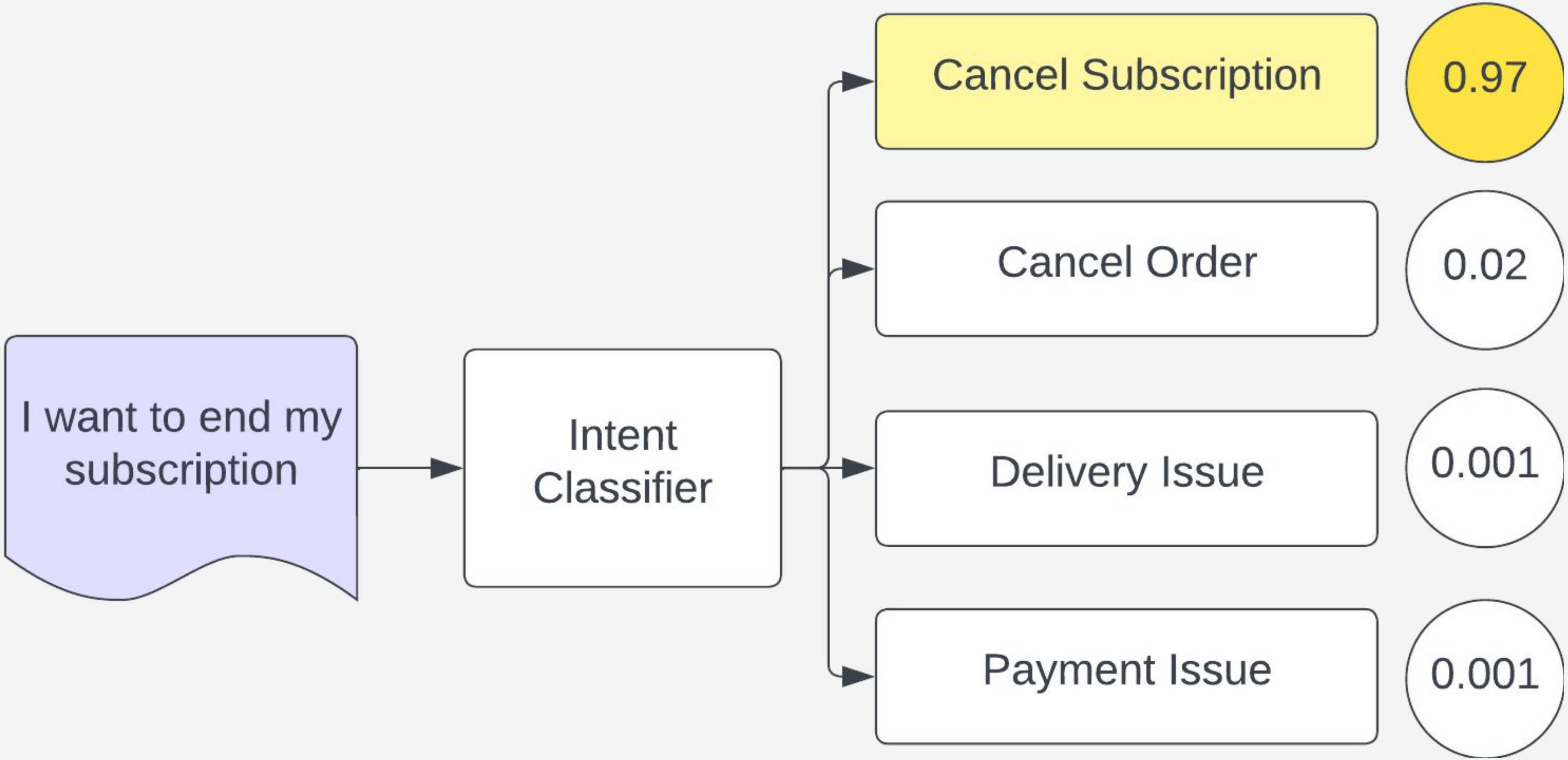
Train Model using SetFit



Why Use SetFit?

- Representation Learning
  - Capture meaningful relationships among items within sets
- Contrastive Objective Function
  - Positive pairs are pulled closer, while negative pairs are pushed apart
- Scalable
- Fast
- SetFit 1.0 was released recently with updated documentation and functionality - thanks to Tom Aarsen!
- No prompting required
- High performance achieved with few examples
- Elegant
- Supported by HuggingFace

Supervised Text Classifier



Batch Inference

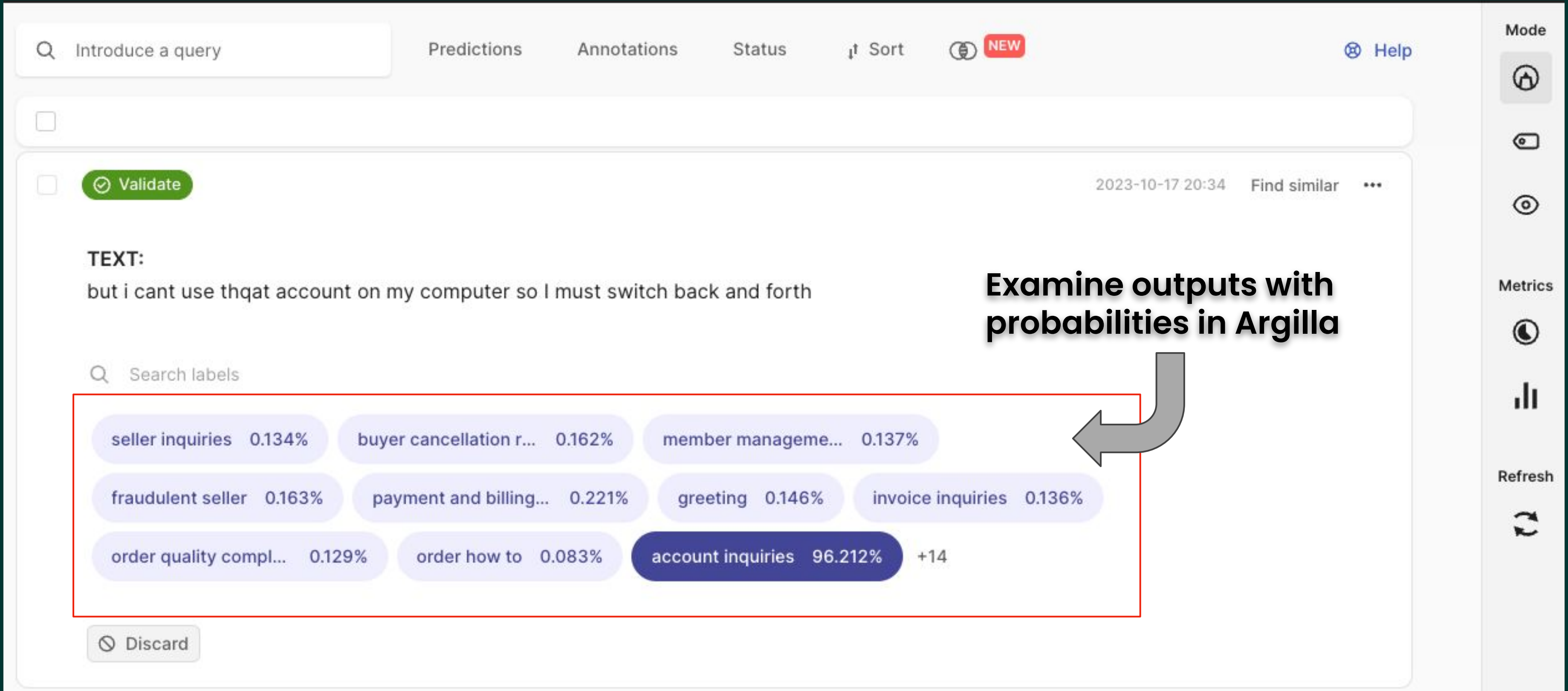
- Run on hold-out dataset to examine classifier on the message-level
    - Create classification report
    - Create confusion matrix
- Error Analysis
    - Examine outputs with probabilities in Argilla

If model version passes message-level criteria:

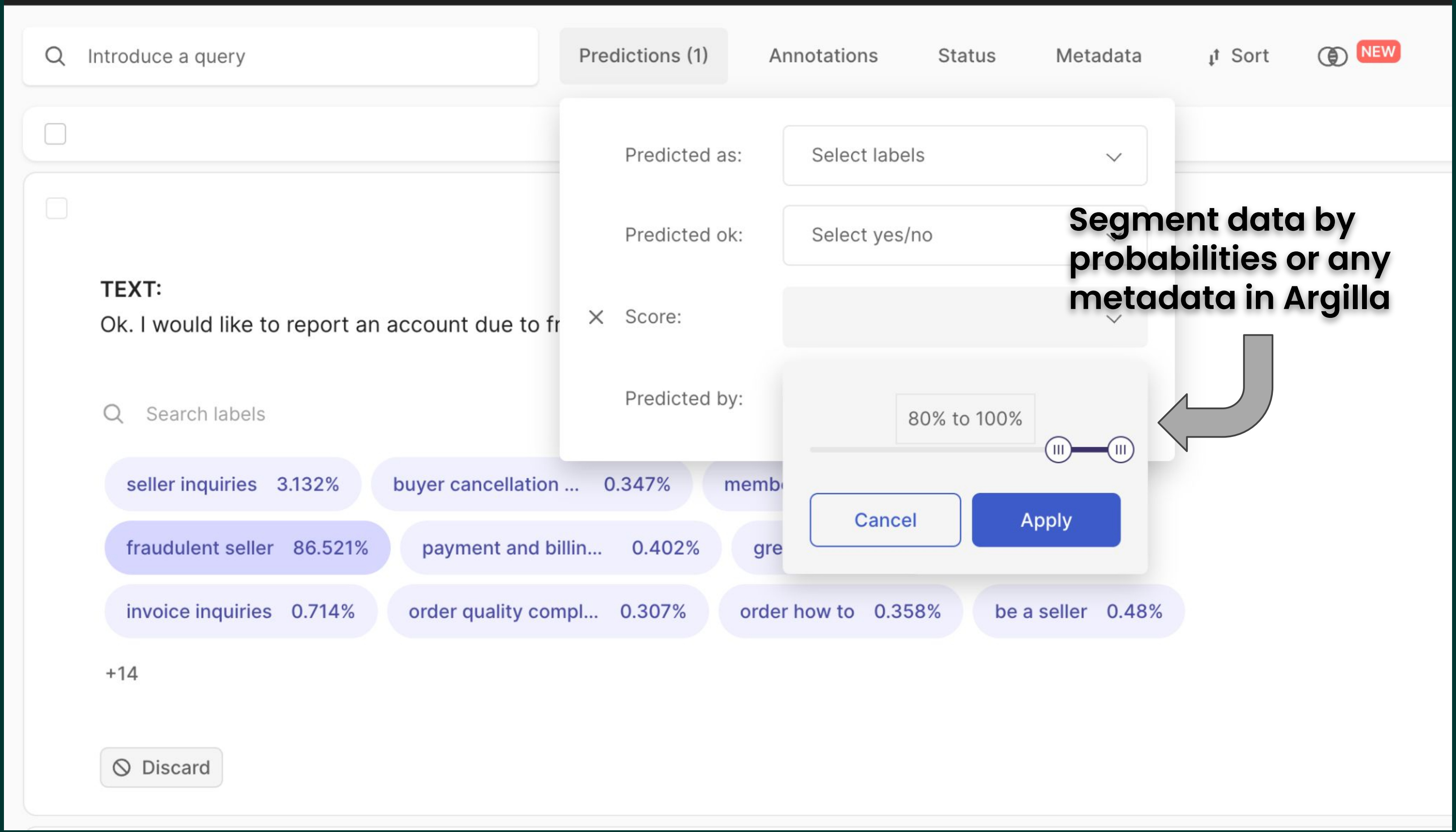
- Run on unseen conversation data to examine classifier on conversation-level

Loris

Examine outputs with probabilities in Argilla



Segment data by probabilities or any metadata in Argilla



Loris

Evaluation

Evaluation Reminder

Coverage: What percent of conversations are assigned an intent

Precision: When an intent is predicted, is it correct?

Frequency: Does the frequency of each contact driver make sense?

Must be evaluated on both the message-level and conversation-level

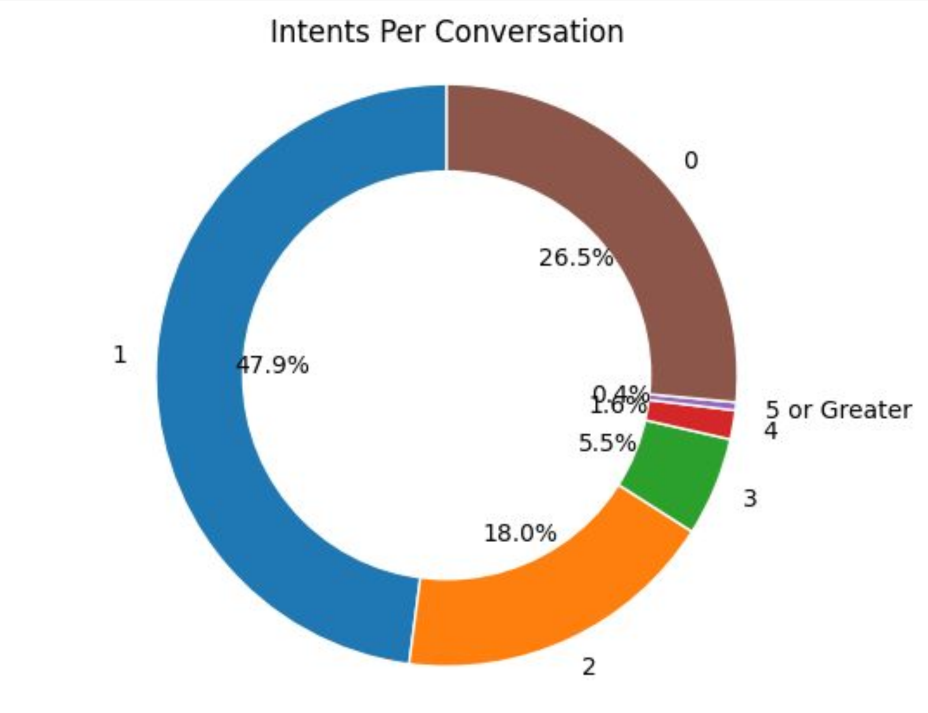
Challenge

How do you evaluate a model on unseen data if you don’t have labels?

- Look at aggregate metrics
- Use human annotators
- Use an LLM

Aggregate Metrics

- Coverage
- Intents per Conversation
- Class Frequency



LLM Validator

- A generative system that can be used to validate training data and validate predictions

Prompt Example: *"Given {class_name} is defined as {definition}, with the following {examples} classify the following message as YES if it is a {class_name} and NO otherwise:"*

- `class_name` is the accepted class name
- `definition` is the accepted class definition
- `examples` are examples for that topic

Loris

Thresholding

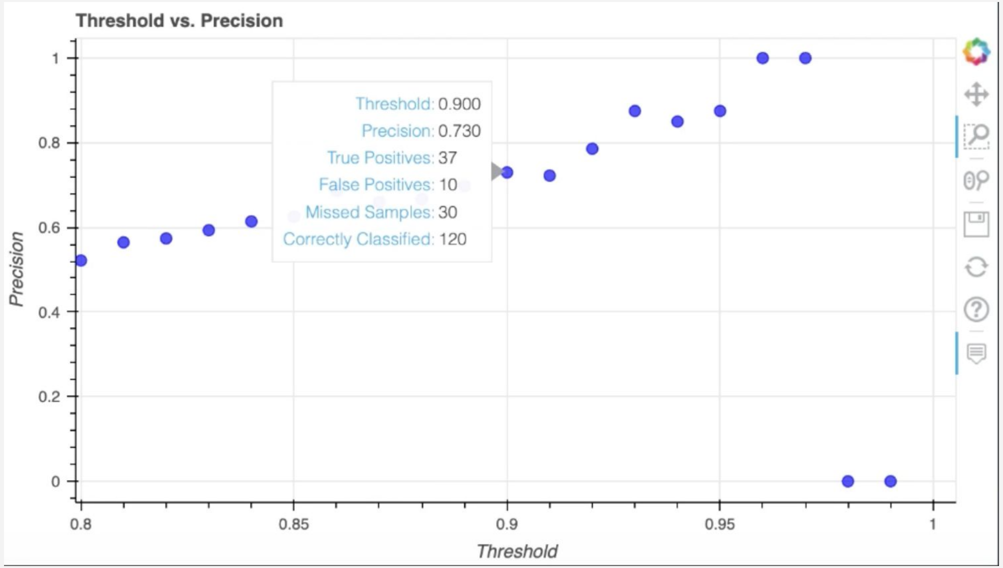
Loris

Examining Threshold

- Text - Sample
- Predicted Class
- Predicted Probability
- Label

The Art of Thresholding

- Simply maximizing some metric or function does not work in practice
- All errors are not equal
- Get your hands dirty
- Examine:
  - True Positives
  - False Positives
  - False Negatives
- Set threshold above where errors get flagrant



Loris

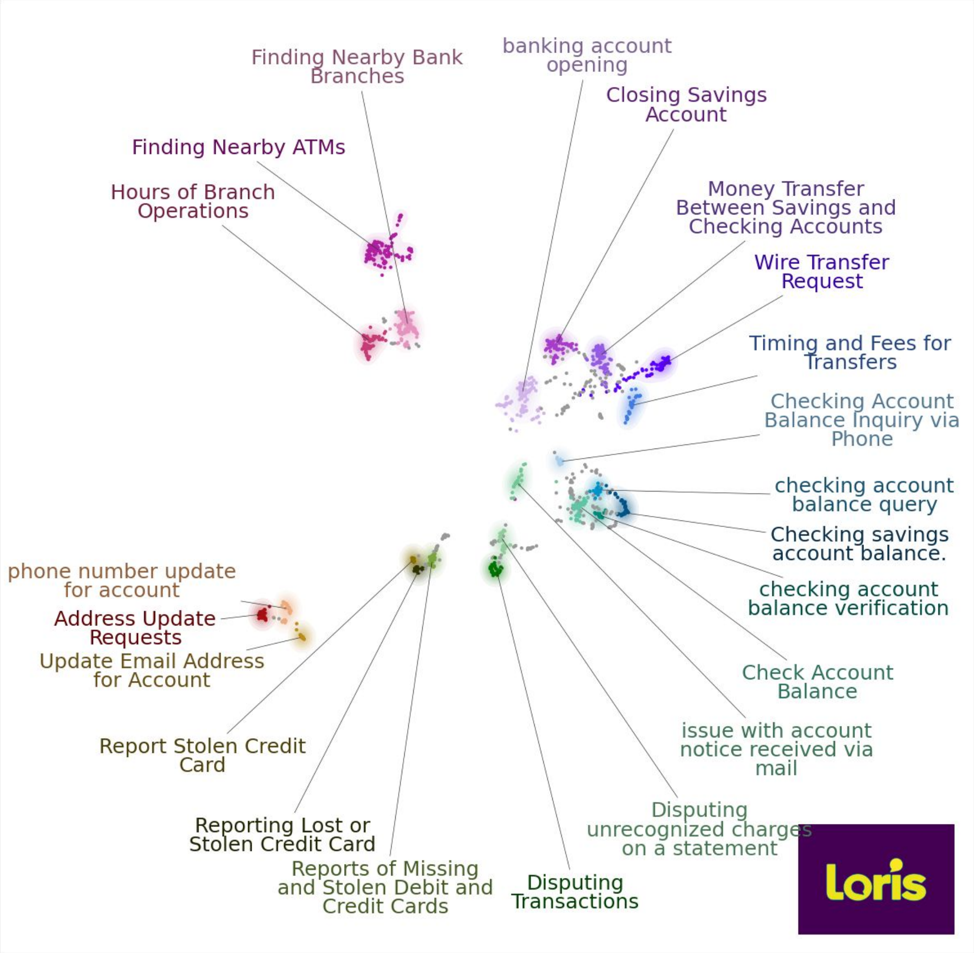
Demo

Demo

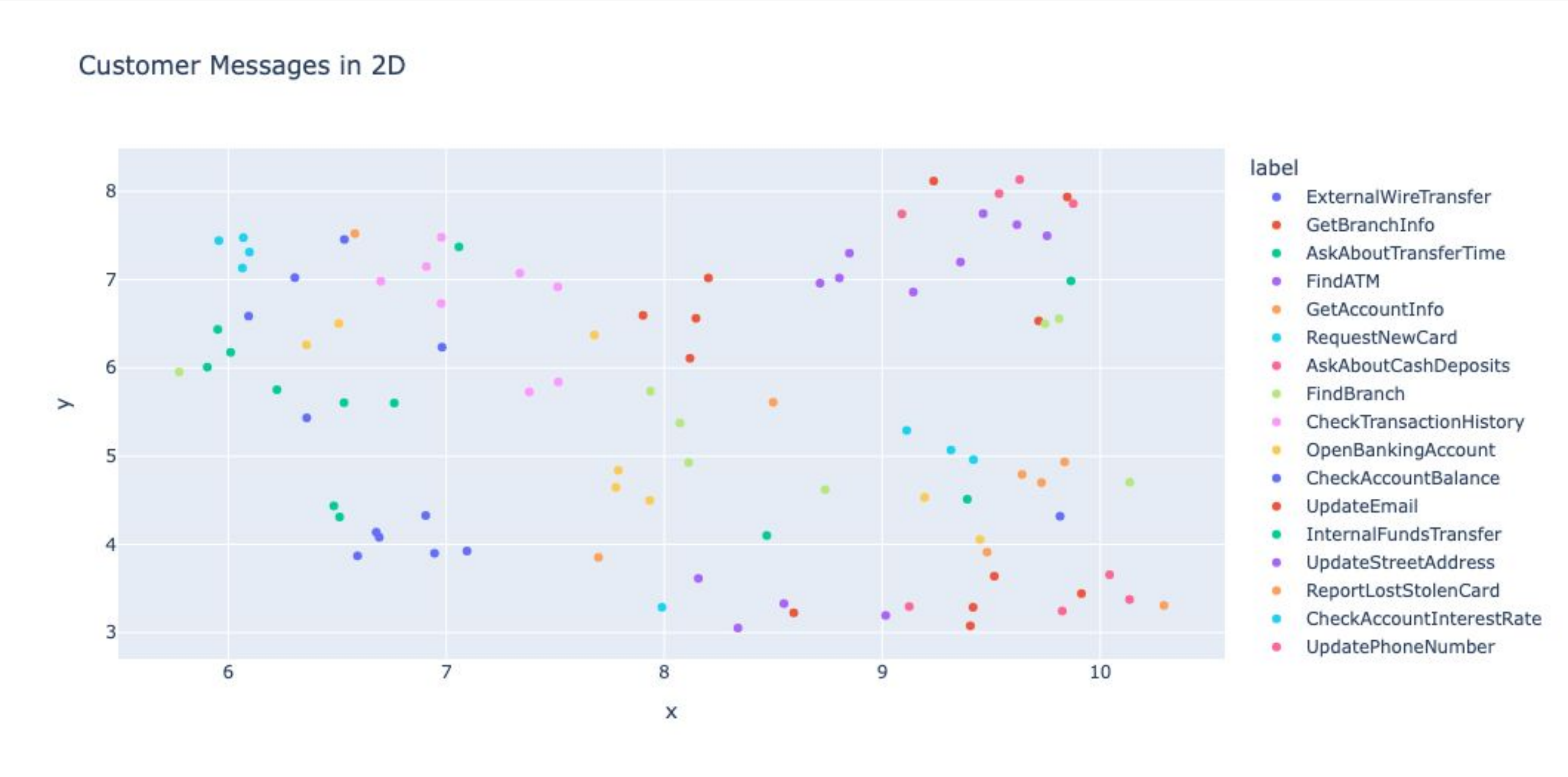
- Dataset: Dialogue State Tracking Challenge from ACL 2023
    - Banking Data
- Explore Data
- Run BERTopic
    - Simulate Intent Discovery
- Train Setfit Model
- Evaluate Classifier
- Visualize what's happening
- Demo Notebooks:
    - Topic Modeling using BERTopic
    - SetFit Train Classifier

Demo - Loris

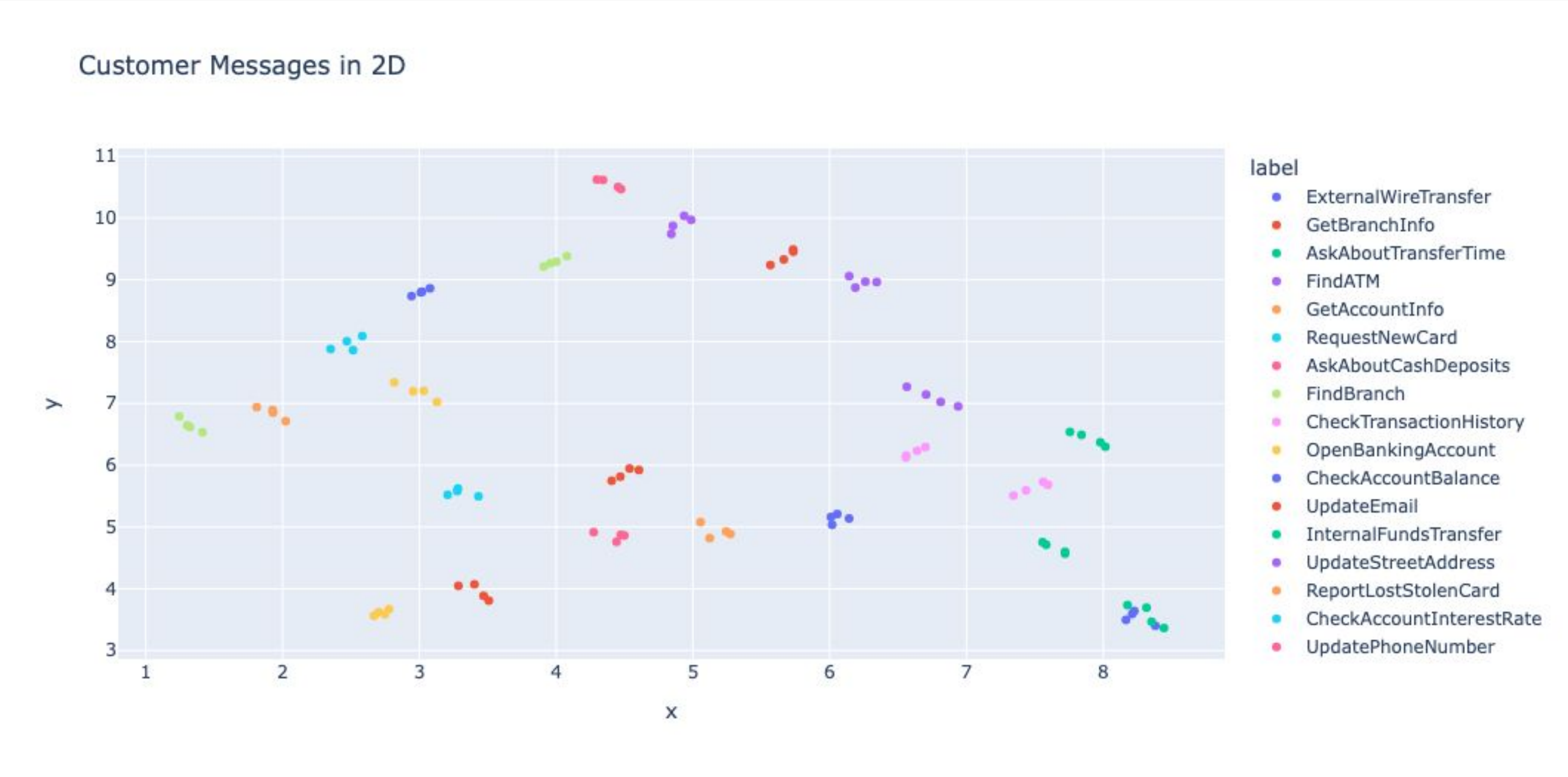
Topics labeled with `KeyBERTInspired + OpenAI GPT-3.5`



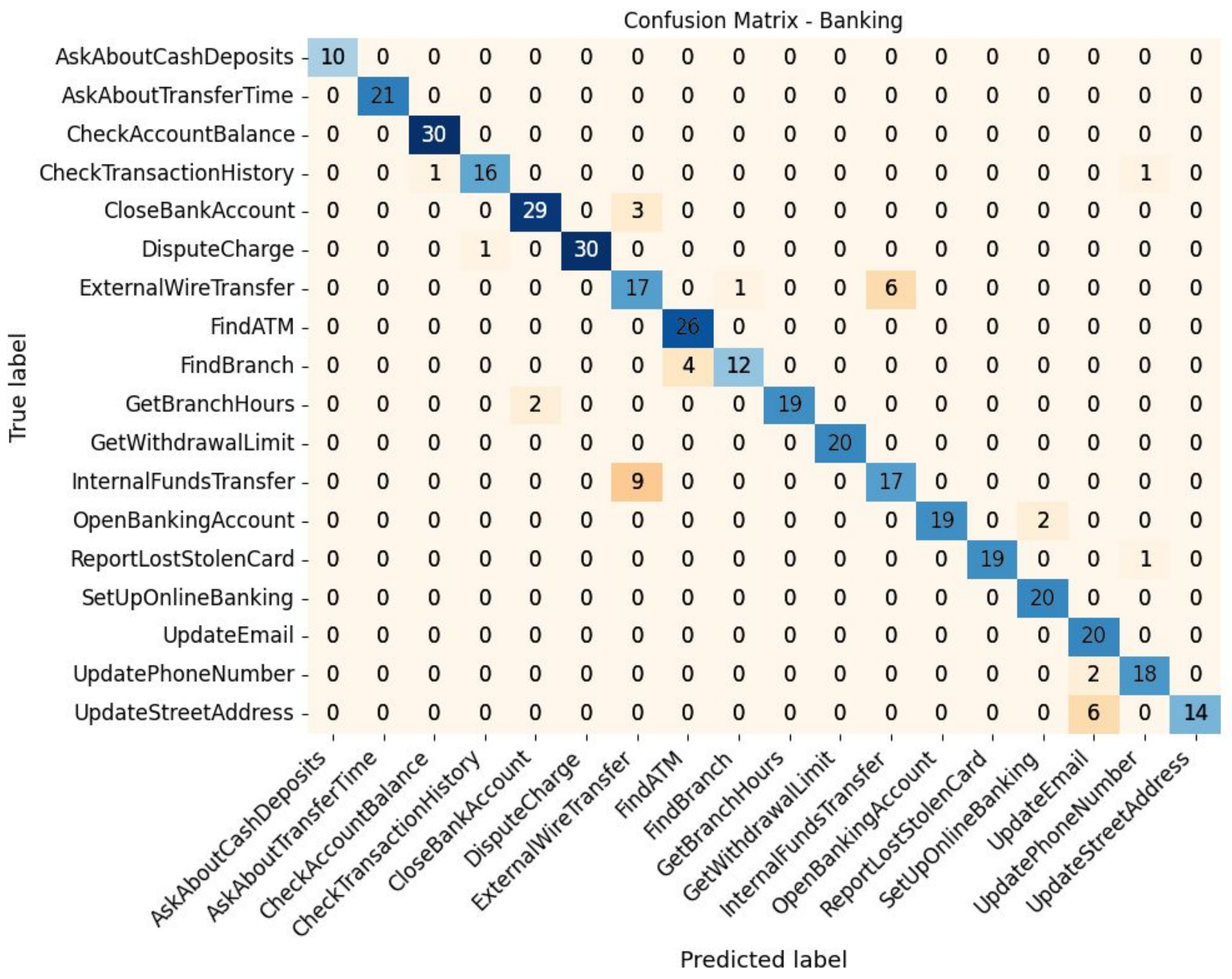
Training Data Before Fine-Tuning



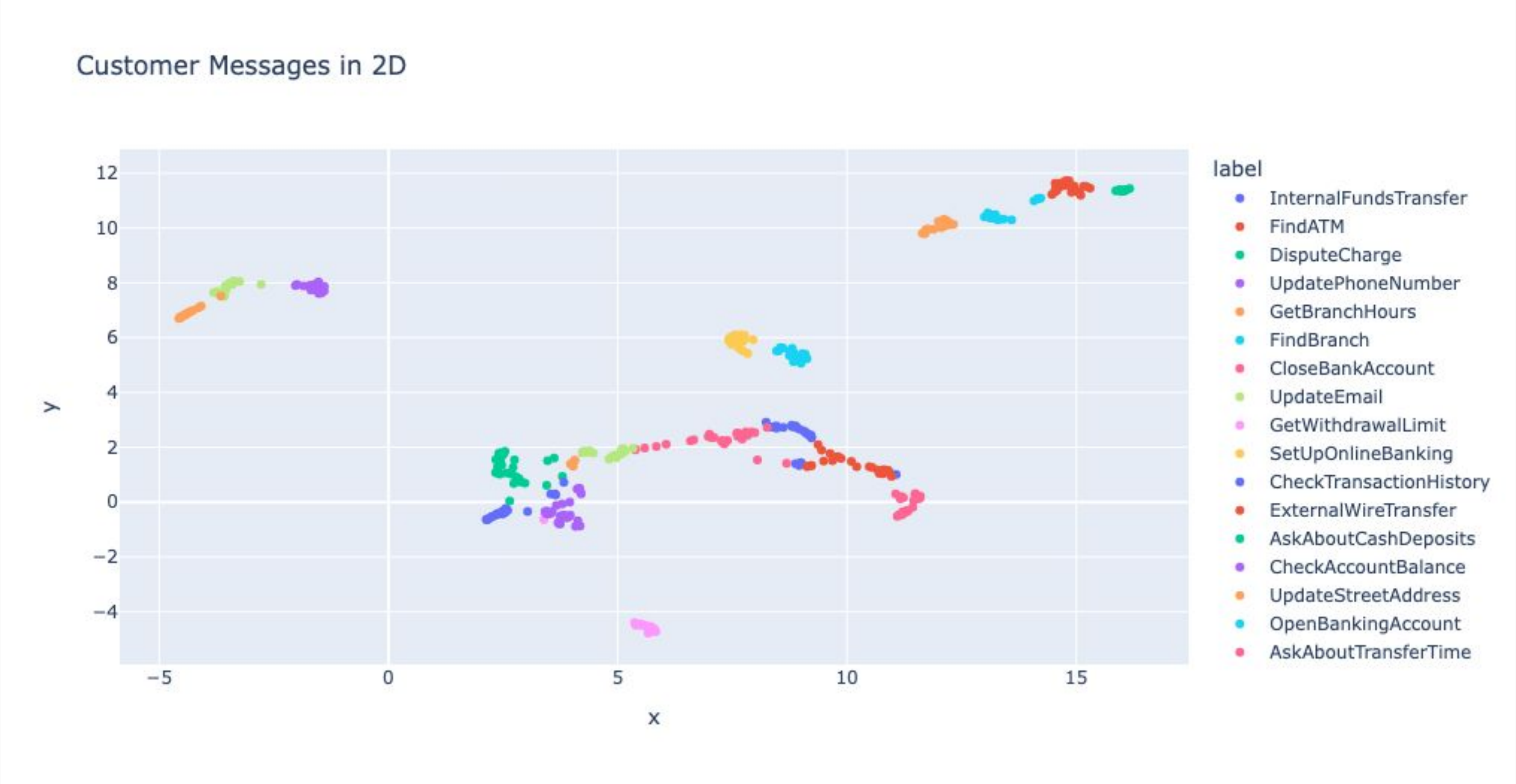
Training Data After Fine-Tuning



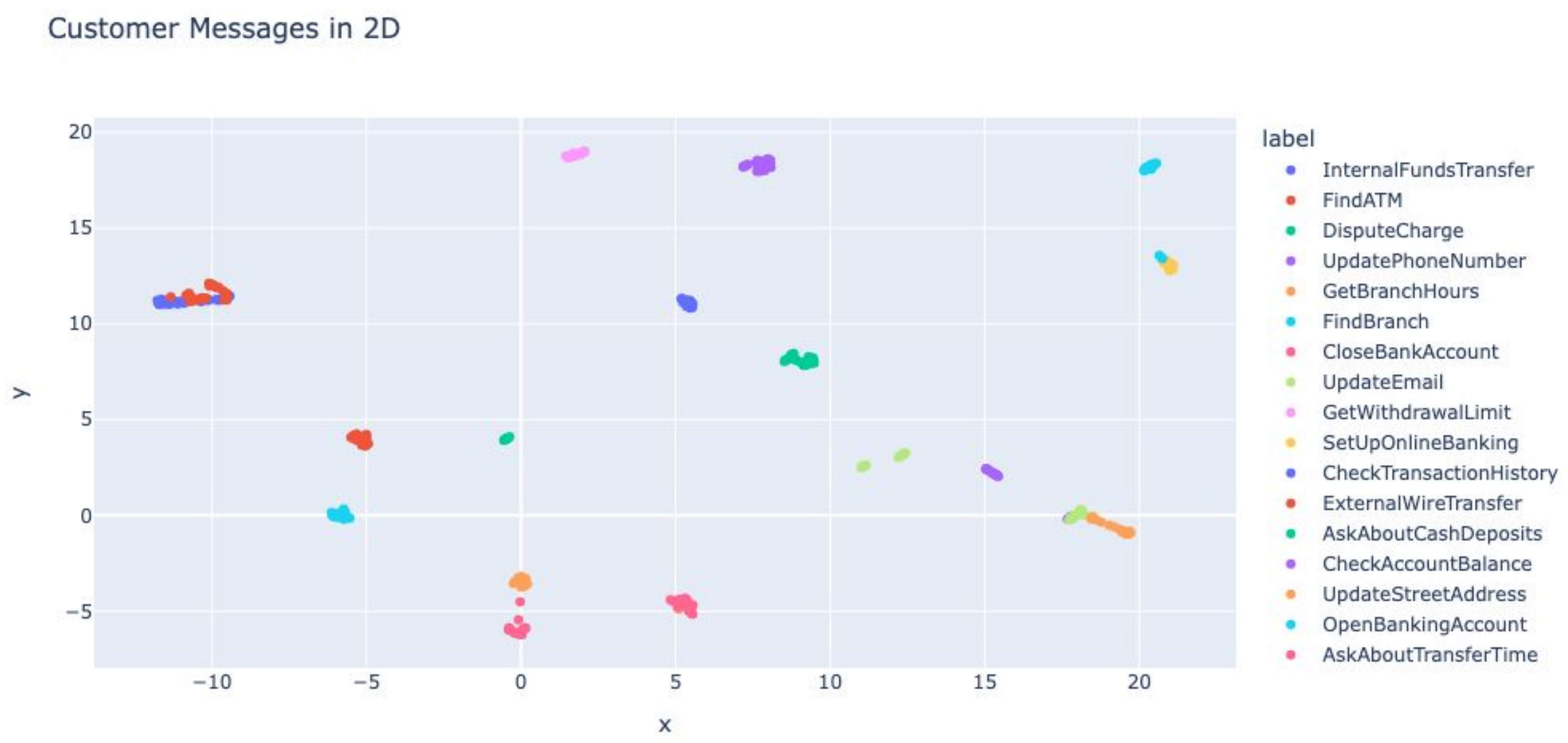
Evaluation


Confusion Matrix - Banking
True label
Predicted label
AskAboutCashDeposits 10 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
AskAboutTransferTime 0 21 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
CheckAccountBalance 0 0 30 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
CheckTransactionHistory 0 0 1 16 0 0 0 0 0 0 0 0 0 0 0 0 1 0
CloseBankAccount 0 0 0 0 29 0 3 0 0 0 0 0 0 0 0 0 0 0
DisputeCharge 0 0 0 1 0 30 0 0 0 0 0 0 0 0 0 0 0 0
ExternalWireTransfer 0 0 0 0 0 0 17 0 1 0 0 6 0 0 0 0 0 0
FindATM 0 0 0 0 0 0 0 26 0 0 0 0 0 0 0 0 0 0
FindBranch 0 0 0 0 0 0 0 4 12 0 0 0 0 0 0 0 0 0
GetBranchHours 0 0 0 0 2 0 0 0 0 19 0 0 0 0 0 0 0 0
GetWithdrawalLimit 0 0 0 0 0 0 0 0 0 0 20 0 0 0 0 0 0 0
InternalFundsTransfer 0 0 0 0 0 0 9 0 0 0 0 17 0 0 0 0 0 0
OpenBankingAccount 0 0 0 0 0 0 0 0 0 0 0 0 19 0 2 0 0 0
ReportLostStolenCard 0 0 0 0 0 0 0 0 0 0 0 0 0 19 0 0 1 0
SetUpOnlineBanking 0 0 0 0 0 0 0 0 0 0 0 0 0 0 20 0 0 0
UpdateEmail 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 20 0 0
UpdatePhoneNumber 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 2 18 0
UpdateStreetAddress 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 6 0 14
AskAboutCashDeposits
AskAboutTransferTime
CheckAccountBalance
CheckTransactionHistory
CloseBankAccount
DisputeCharge
ExternalWireTransfer
FindATM
FindBranch
GetBranchHours
GetWithdrawalLimit
InternalFundsTransfer
OpenBankingAccount
ReportLostStolenCard
SetUpOnlineBanking
UpdateEmail
UpdatePhoneNumber
UpdateStreetAddress

Evaluation Data Before Fine-Tuning



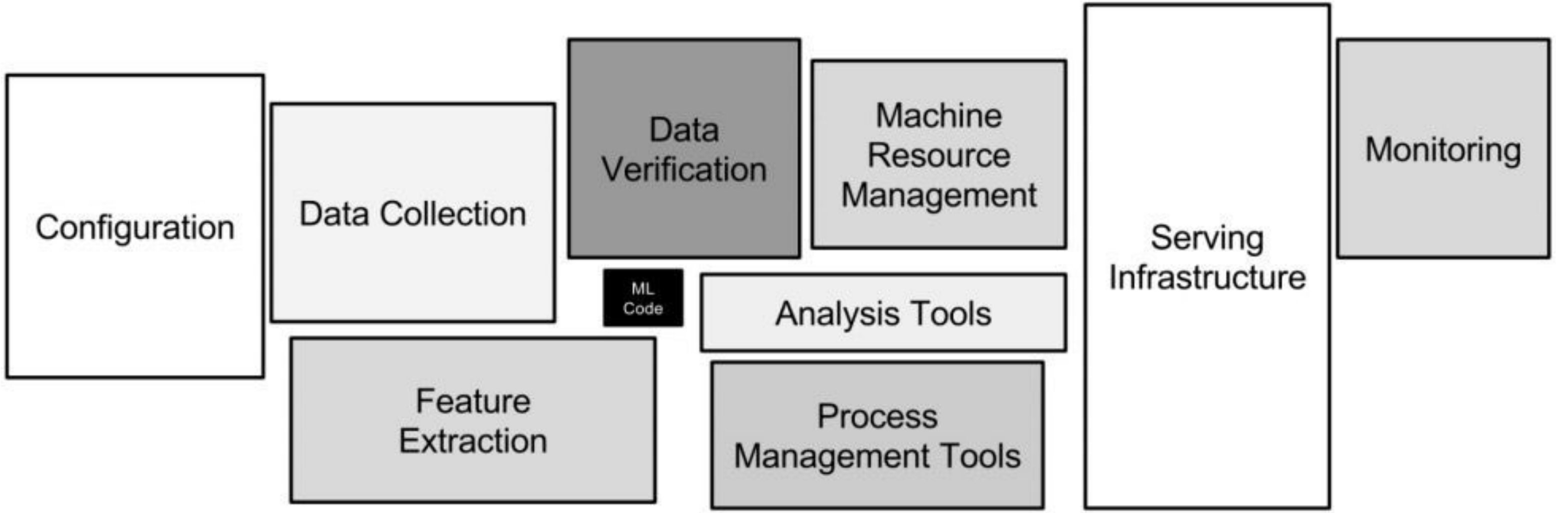
Evaluation Data After Fine-Tuning



Conclusion

Hidden Technical Debt in Machine Learning Systems

D. Sculley, Gary Holt, Daniel Golovin, Eugene Davydov, Todd Phillips
{dsculley,gholt,dgg,edavydov,toddphillips}@google.com
Google, Inc.



Learnings

- Understand the problem and use-case deeply
    - Keep revisiting the problem and underlying need
- Use the learnings from other models
    - Create sanity checks - eyeball test sets
- Incorporating classes to ignore improves performance
- Incorporating heuristics can really help
- Training a model is small part
- Creating high quality datasets can help you long term

Learnings

- Segmenting data is a must
- Threshold tuning is almost always needed for multi-class problems
- Creating strong components is important
- You're creating a system not a model
    - Ability to modify is a requirement
- Automation and end to end is a great motivation, but know your data
    - Examine, visualize, explore
    - Include human in the loop when necessary

Argilla

DataMapPlot

W&B

SBERT.net

Sentence-Transformers

spaCy

BERTopic

PHOENIX

Takeaways

- Introduction to Loris
- Creating systems to help companies solve problems
  - Challenges in creating client-specific models
  - Example of how to apply NLP to solve problems
- BERTopic is a one-stop shop for topic modeling
- LLMs are great, but they are one potential solution, not the only solution
  - If you're creating a classifier, just create a classifier :)
- Combining multiple approaches often yields the best results

Q & A

Thank You!

Learning from Machine Learning

Loris.ai
Seth Levine | LinkedIn

References

- Loris.ai
- Learning from Machine Learning - YouTube
- BERTopic
- Argilla docs
- Sentence-Transformers
- SetFit
- DataMapPlot documentation
- Hidden Technical Debt in Machine Learning Systems
- It’s the Golden Age of Natural Language Processing, So Why Can’t Chatbots Solve More Problems?
- Getting Machine Learning Projects from Idea to Execution

Loris

Learning from Machine Learning

Learning from Machine Learning

Loris CoPilot

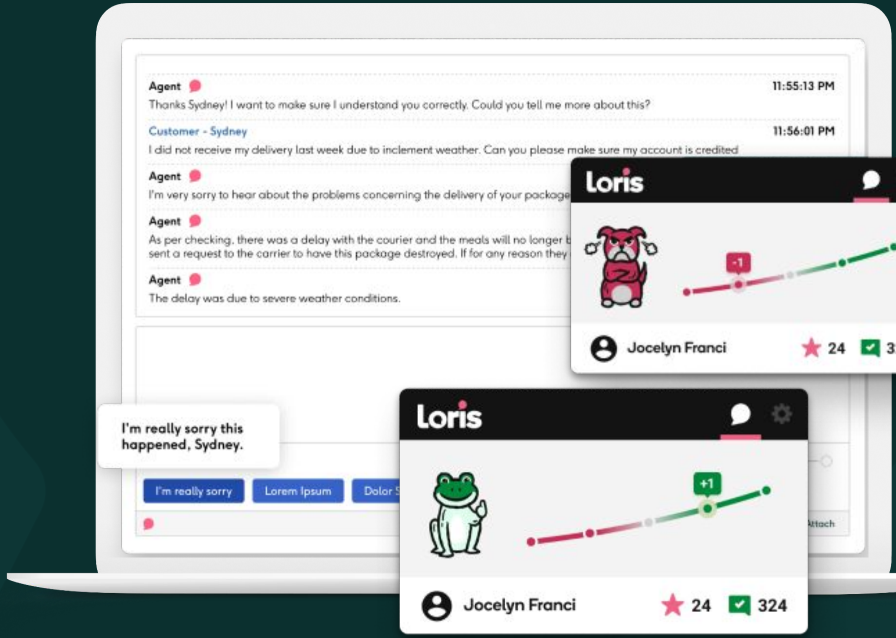
Real-time Agent Assistance

Live agent guidance & feedback

- Uses natural language AI to categorize customer conversations and identify contact drivers
- Provides CSAT score prediction
- Pinpoints both common and emerging areas of friction

Gives leaders

- Single, objective CX benchmarks
- Complete transparency into issues
- Ability to make better use of existing data
- Power to improve CX across multiple functions: product, marketing, customer service, etc.



Loris CQA

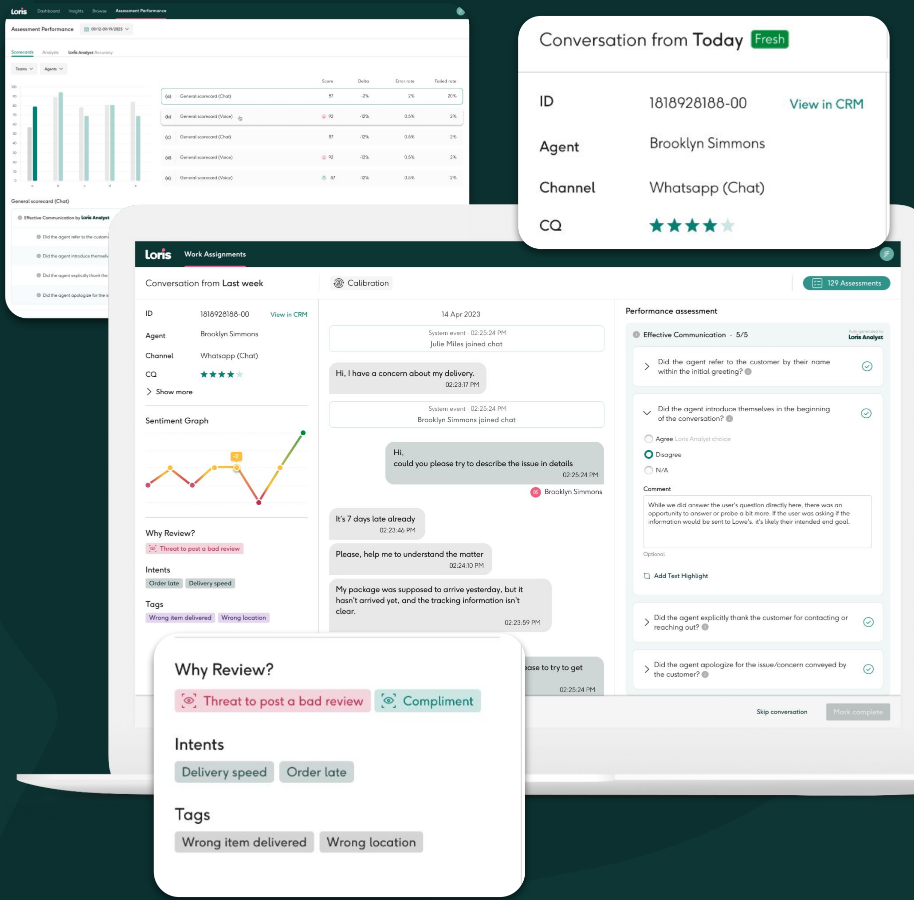
Conversational Quality Assurance

Agent quality & coaching solution

- Analyzes all your customer interactions
- Surfaces most recent & relevant conversations for review
- Automatically scores agent conversation skills
- Provides agent performance & coaching summaries

Helps QA and Agent Managers

- Lower the cost of QA
- Improve agent performance in less time
- Increase consistency & value in QA



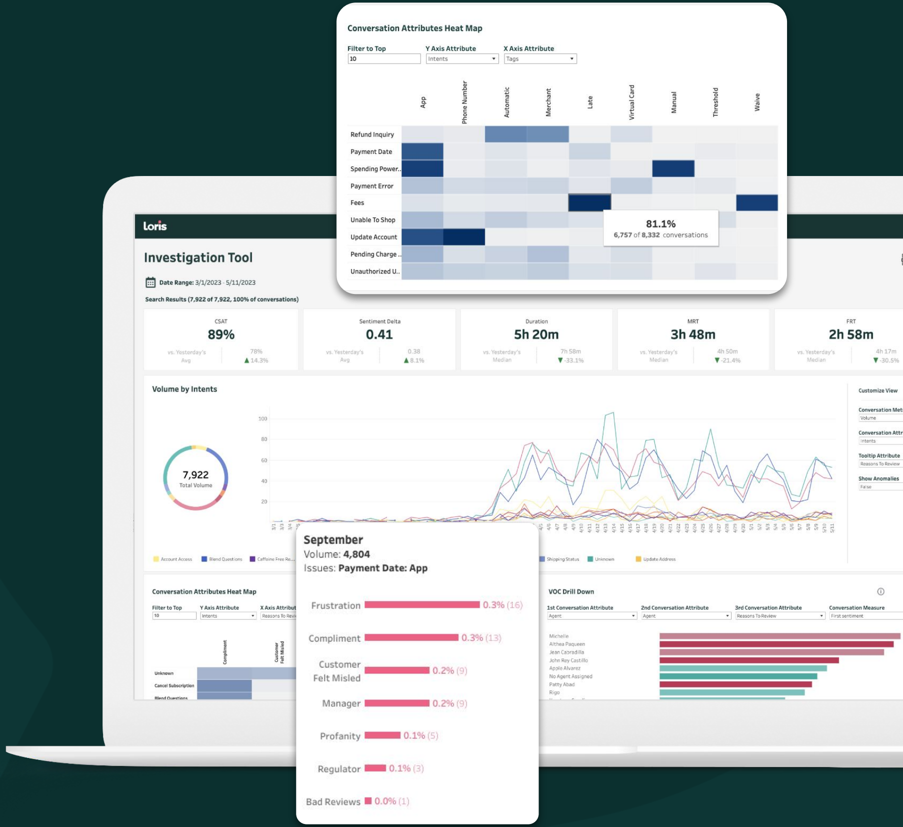
Loris Insights
Customer interaction analytics

CX intelligence platform

- Uses natural language AI to categorize customer conversations and identify contact drivers
- Provides CSAT score prediction
- Pinpoints both common and emerging areas of friction

Gives leaders

- Single, objective CX benchmarks
- Complete transparency into issues
- Ability to make better use of existing data
- Power to improve CX across multiple functions: product, marketing, customer service, etc.



Loris